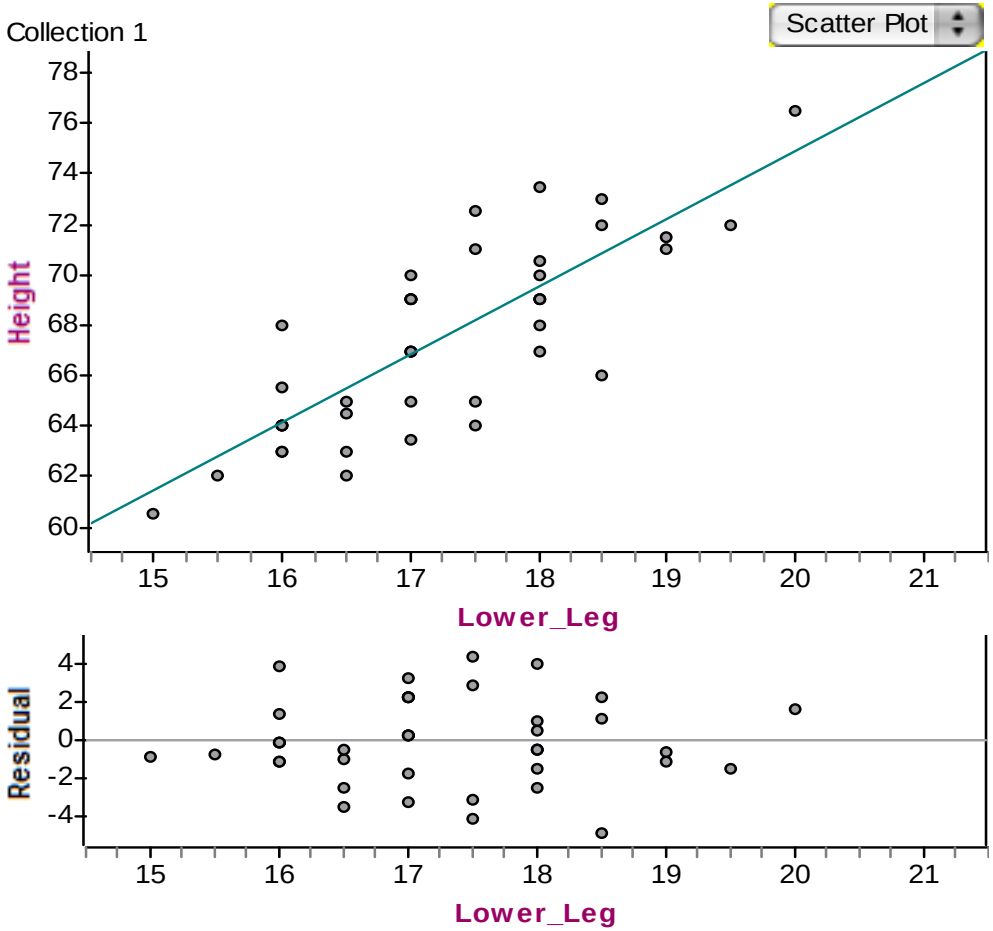
PREDICTING HEIGHTS PROJECT

INTRODUCTION

- Our group measured…
  - Arm length- From shoulder to farthest finger tip.
  - Wrist circumference- Wrapped tape measure around the subject's wrist.
  - Lower Leg- (Had subject sit down) From top of knee to ankle.
  - Actual height and gender were also collected.

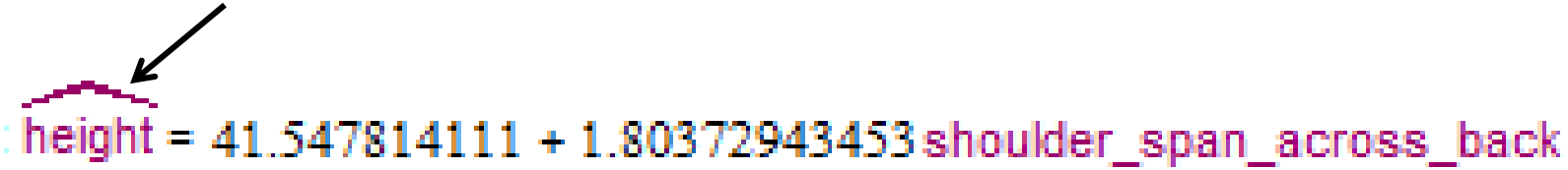
HEIGHT VS. LOWER LEG



HEIGHT VS. LOWER LEG

- Describe graph (form, direction, strength)
- Give correlation value (r)

LSR line



- Interpret slope…
- Interpret r-squared…
- Describe residual plot (form, direction, strength)
- Is the linear model appropriate? Use correlation, residual plot, and original plot

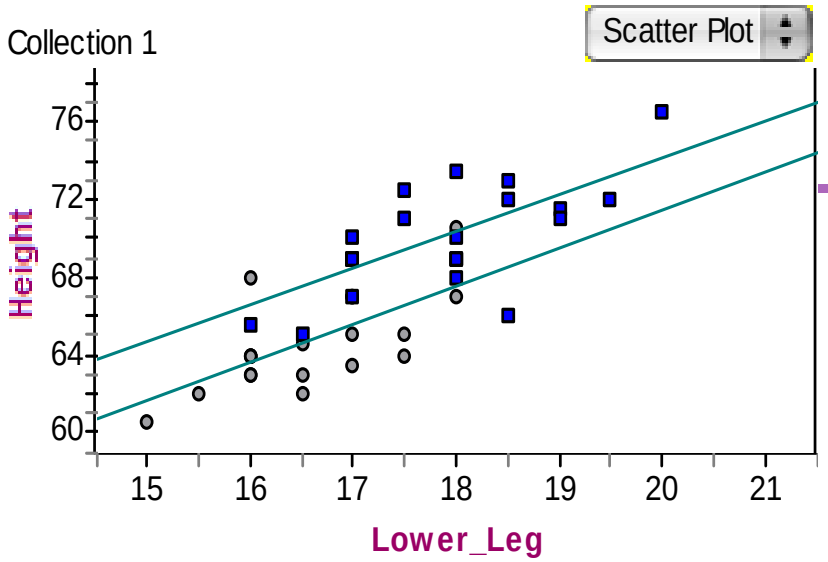
HEIGHT AND LOWER LEG BASED ON GENDER

FEMALES:

- Describe female data
- List Female LSR line
- List female correlation
- List female r-squared

MALES:

- Describe male data
- List male LSR line
- List male correlation
- List male r-squared

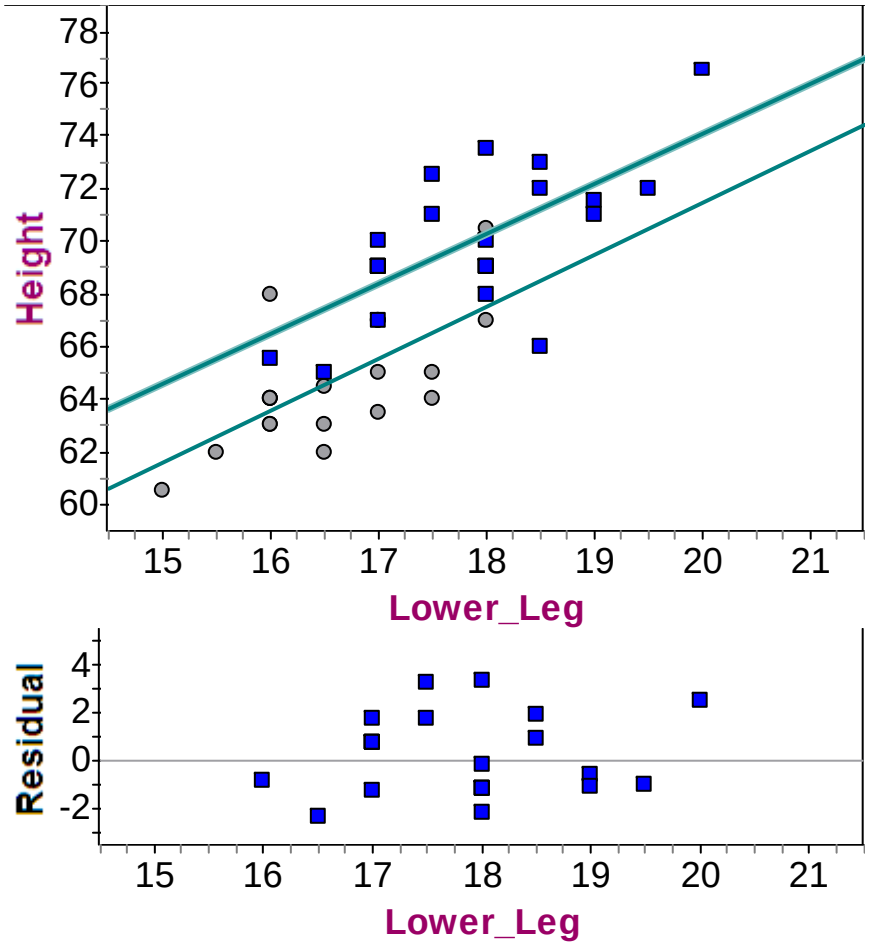


Compare male and female data

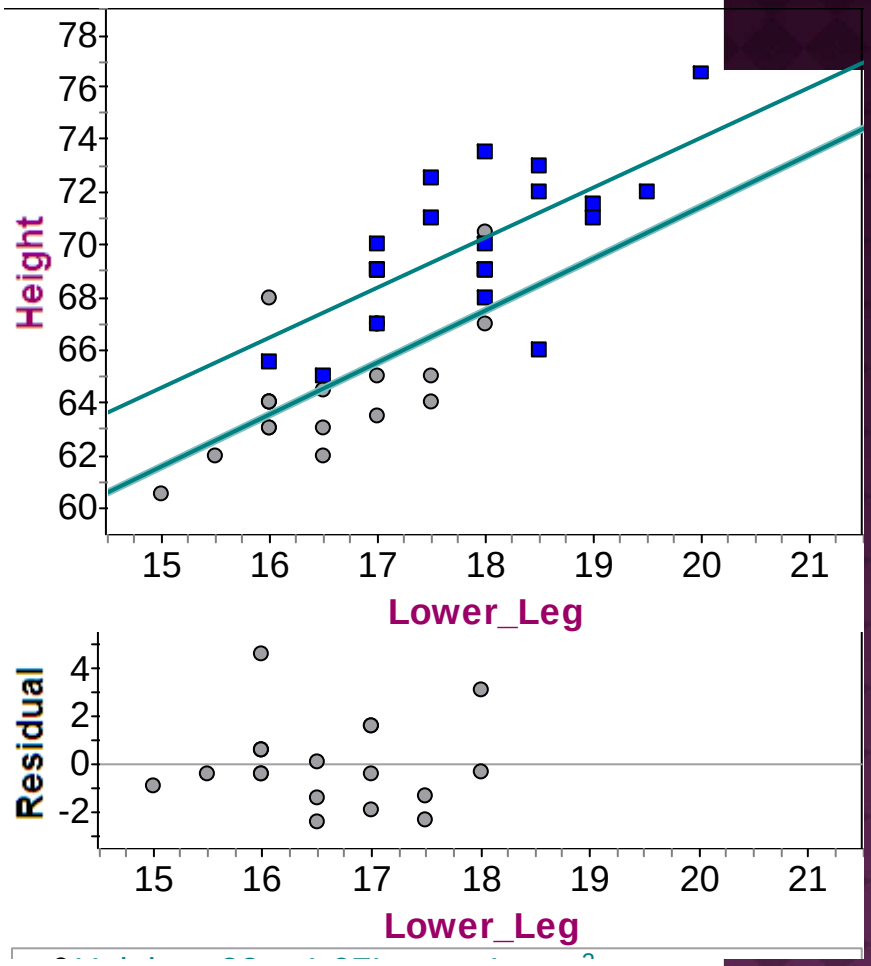
male_height = 49.4989155547 + 1.32094314587shoulder_span_across_back

female_height = 42.5629116326 + 1.66388640364shoulder_span_across_back

MALE RESID PLOT

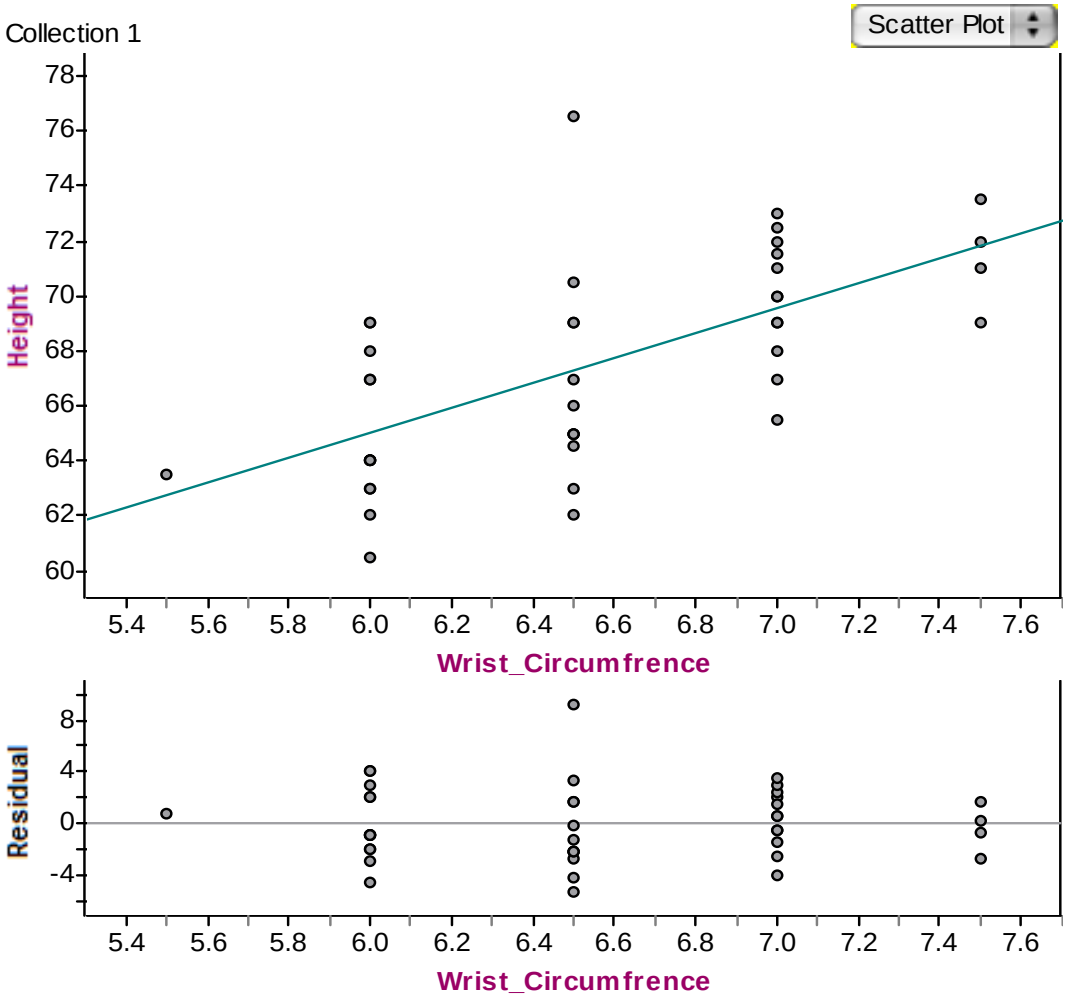


FEMALE RESID PLOT



Comment on the fit of the linear model for each gender

HEIGHT VS. WRIST CIRCUMFERENCE



HEIGHT VS. WRIST CIRCUMFERENCE

- Describe graph (form, direction, strength)
- Give correlation value (r)

- HEIGHT = a + b(wrist circ)

- Interpret slope…
- Interpret r-squared…
- Describe residual plot (form, direction, strength)
- Is the linear model appropriate? Use correlation, residual plot, and original plot

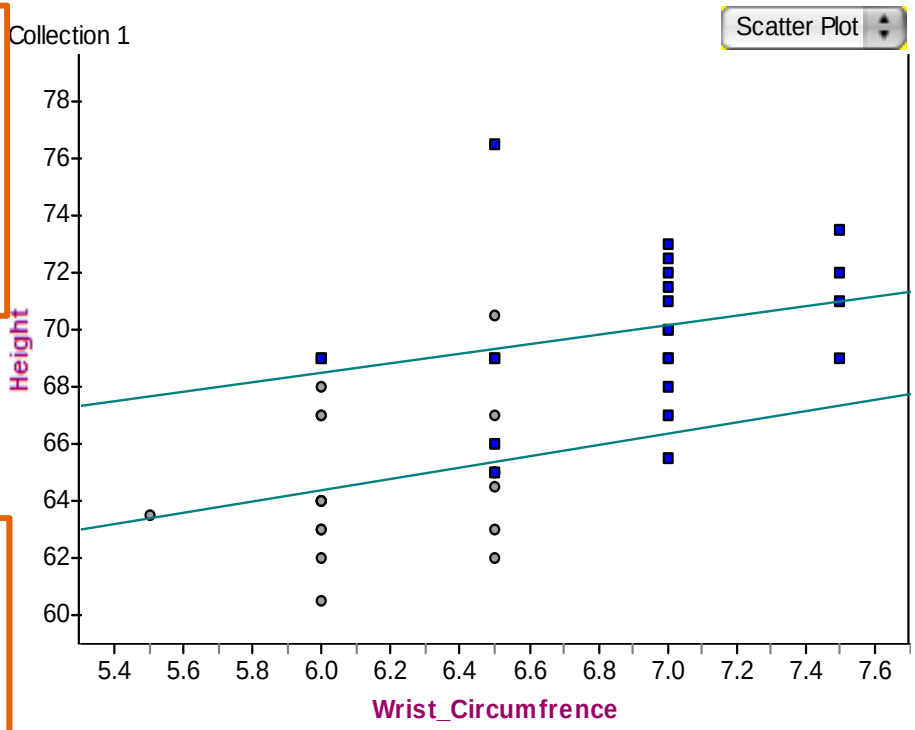
HEIGHT AND WRIST CIRCUMFERENCE BASED ON GENDER

FEMALES:

- Describe female data
- List female LSR line
- List female correlation
- List female r-squared

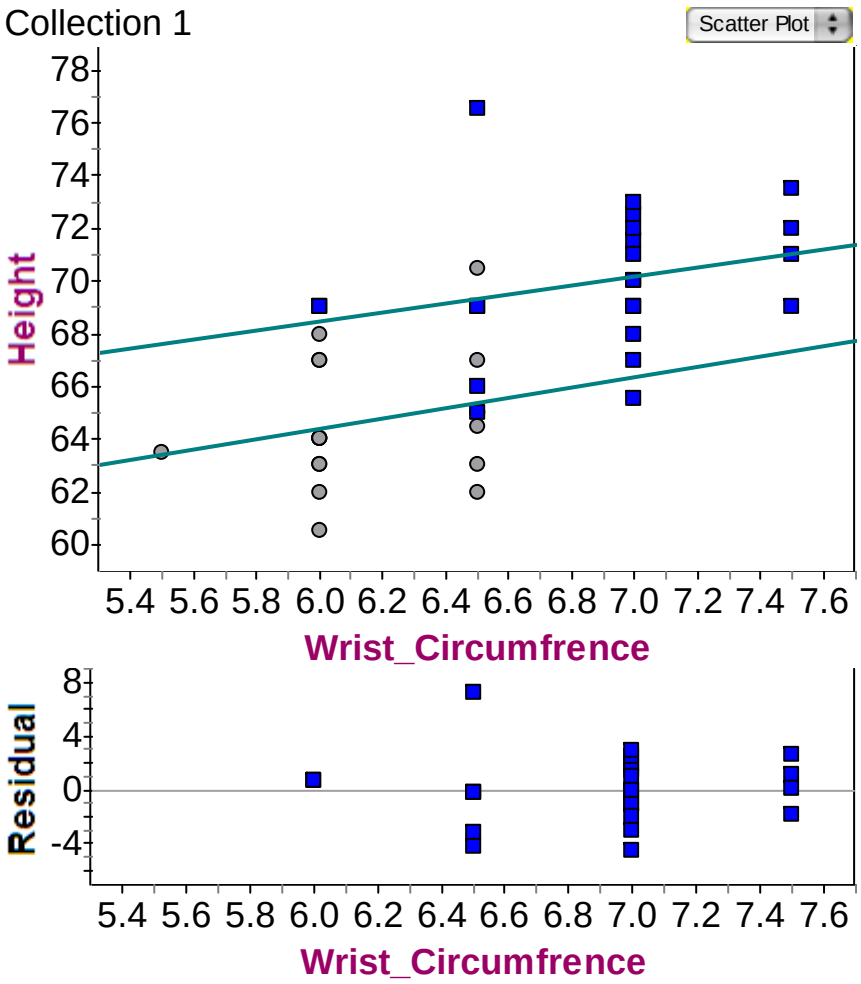
MALES:

- Describe male data
- List male LSR line
- List male correlation
- List male r-squared

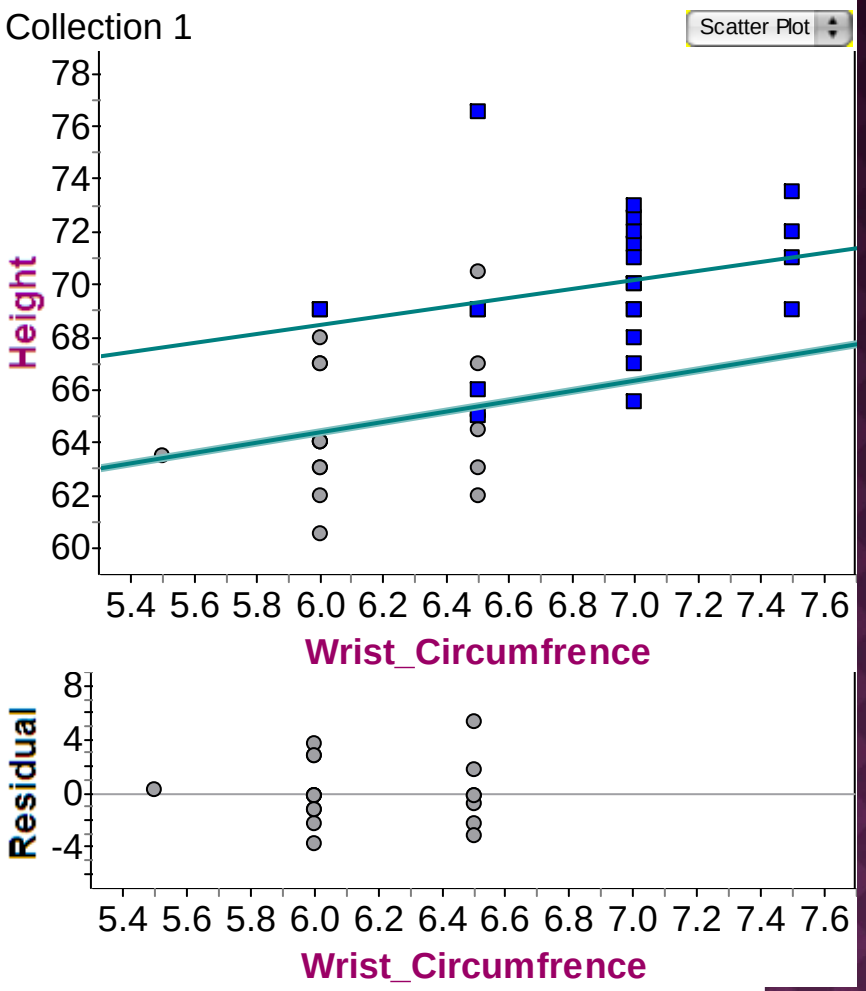


Compare male and female data

MALE RESID PLOT

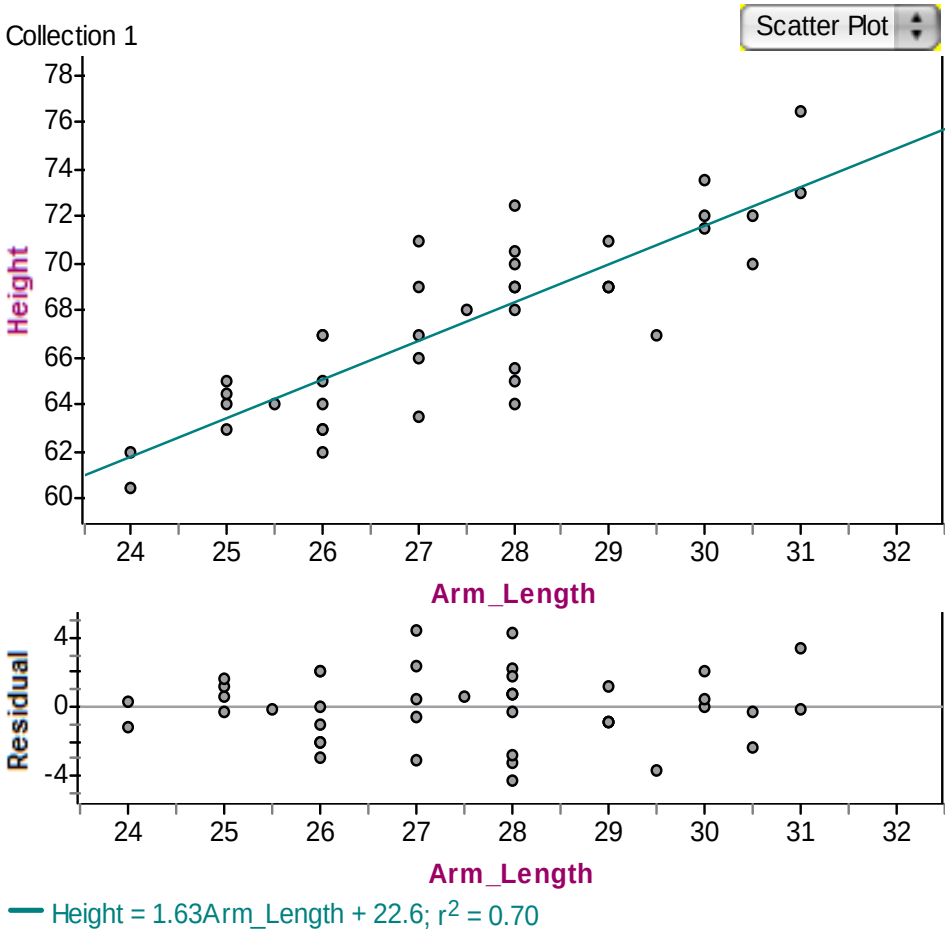


FEMALE RESID PLOT



Comment on the fit of the linear model for each gender

HEIGHT VS. ARM LENGTH



- Describe graph (form, direction, strength)
- Give correlation value (r)

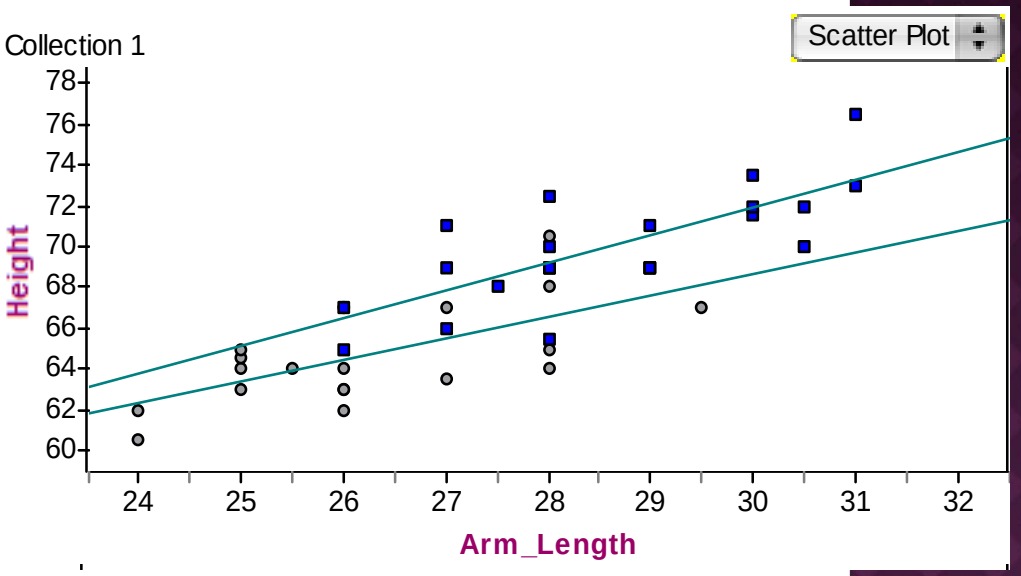
- HEIGHT = a + b(arm length)

- Interpret slope…
- Interpret r-squared…
- Describe residual plot (form, direction, strength)
- Is the linear model appropriate? Use correlation, residual plot, and original plot

HEIGHT AND ARM LENGTH BASED ON GENDER

FEMALES:

- Describe female data
- List female LSR line
- List female correlation
- List female r-squared

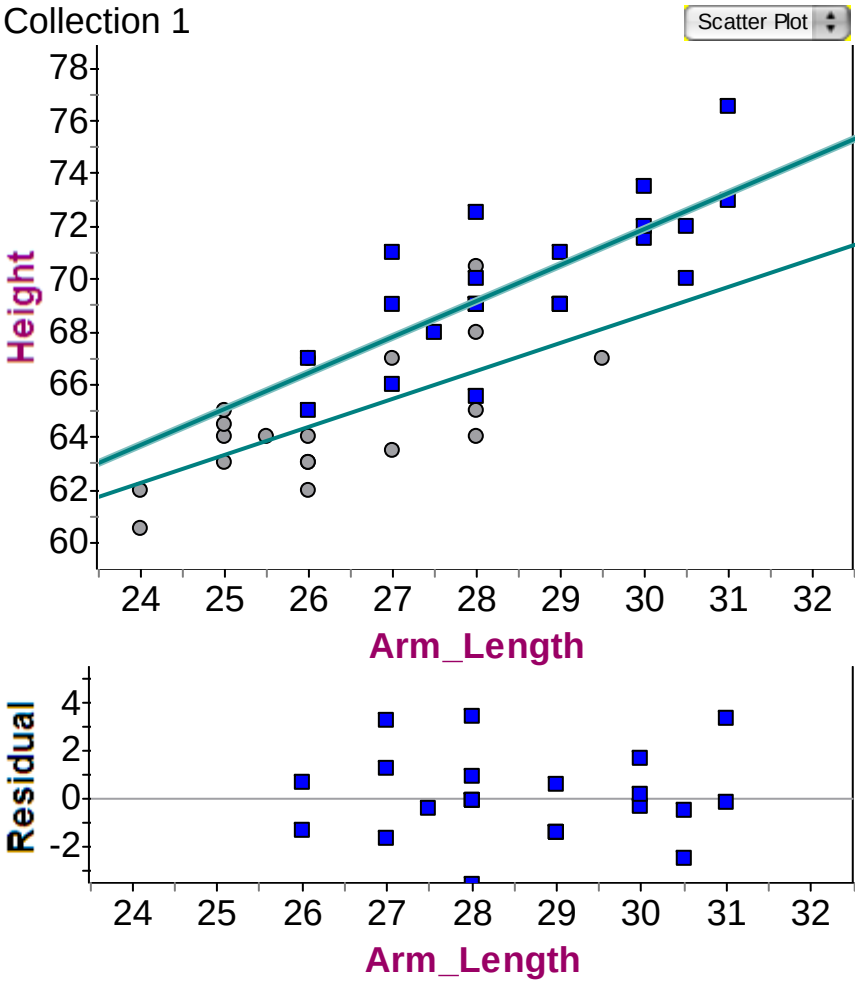


MALES:

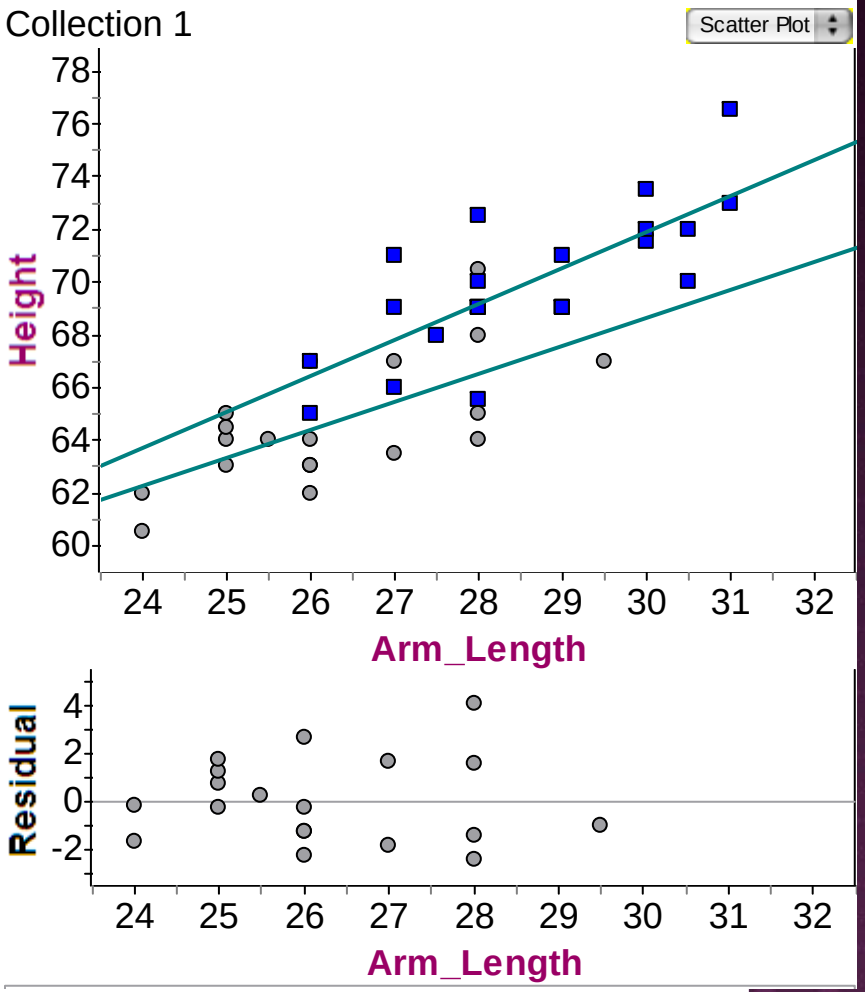
- Describe male data
- List male LSR line
- List male correlation
- List male r-squared

Compare male and female data

MALE RESID PLOT



FEMALE RESID PLOT



Comment on the fit of the linear model for each gender

BEST MODEL – LOWER LEG

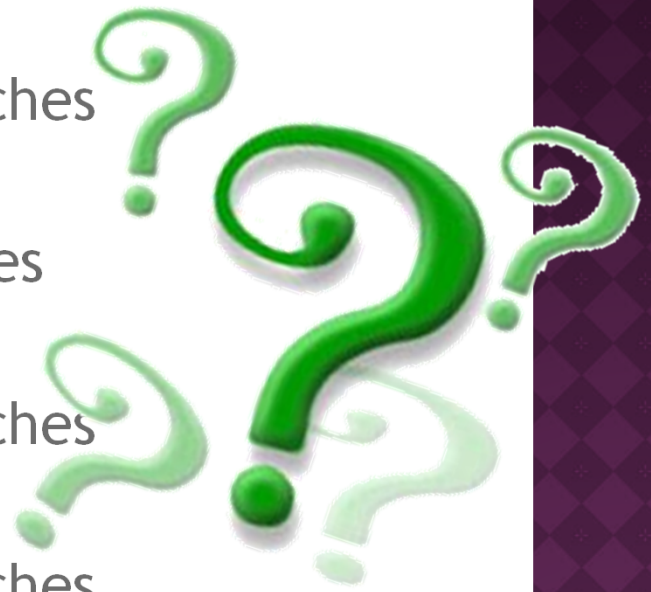
- Lower leg length seems to be the best model used to predict height
- The residual plot is scattered and doesn't appear to have any pattern
- There is a fairly strong correlation of .8, which means it is a moderately strong plot
- 64% of the change in height is due to the change in lower leg
  - This helps to justify that more than 2/3 of the data's height is associated with lower leg length
- Gender doesn't seem to affect this variable and their correlations are very similar
  - .671 for females and .683 for males
  - This model will work equally as well regardless of gender

OUR PREDICTED HEIGHTS

- Partner #1
  - Lower leg measurement = 17.5 inches
  - Height = 21.2 + 2.68(17.5) = 68.1 inches
  - Actual Height = 64 inches
  - Residual = 64 – 68.1 = -4.1 inches
  - Overestimate
- Partner #2
  - Lower leg measurement = 18 inches
  - Height = 21.2 + 2.68(18) = 69.44 inches
  - Actual Height = 67 inches
  - Residual = 67 – 69.44 = -2.44 inches
  - Overestimate
- Partner #3
  - Lower leg measurement = 17.5 inches
  - Height = 21.2 + 2.68(17.5) = 68.1 inches
  - Actual Height = 71 inches
  - Residual = 71 – 68.1 = 2.9
  - Underestimate

PREDICTED TEACHER HEIGHTS

- Mrs. McNelis
  - Height = 2.68(17.5) + 21.2 = 68.1 inches
- Mrs. Ladley
  - Height = 2.68(17.5) + 21.2 = 68.1 inches
- Mr. Smith
  - Height = 2.68(17) + 21.2 = 66.8 inches
- Mrs. Tannous
  - Height = 2.68(17.5) + 21.2 = 68.1 inches
- Miss Gemgnani
  - Height = 2.68(17.5) + 21.2 = 68.1 inches
- Mrs. Bolton
  - Height = 2.68(17) + 21.2 = 66.8 inches



CONFIDENCE

- Based on what we have learned in class we are confident in our predictions
  - Strong correlation, linear original plot, scattered residual plot
- However, based on our predictions we are not as confident
  - Large residuals in our own heights
  - Not all the teachers are the same height
- If we measured more specific (instead of the nearest half inch) we could have obtained a more accurate measurement
  - More confident

LINREG T TEST (ON BEST MODEL)

- Create linear model for best measurement and copy onto ppt

Model of Collection 1 — Multiple Regression

Response attribute (numeric): Height

Predictor	Coefficient	Std Error	t Statistic	P Value	ΔR^2
Constant	21.1544	5.4784	3.861	0.0004	
Lower_Leg	2.6823	0.3165	8.474	0.0000	0.6423

Regression Equation: $\widehat{Height}$ = 21.1543899252 + 2.6823096

R-Squared: **0.642265**
Adjusted R-Squared: **0.633321**
Standard Deviation of the Error: **2.24041**

Do lin Reg t test using the output above

Ho: $\beta_1 = 0$
Ha: $\beta_1 >, <, \neq 0$

Conditions (with graphs!) & statement

$$t = \frac{2.6823}{0.3167} = 8.474$$

$P(t > 8.474 | df = n-2) =$

- We reject Ho….
- We have sufficient evidence…
- Therefore…

LIN REG CONFIDENCE INTERVAL

- Since we rejected Ho, we need to complete a conf. int.

Statement

$b \pm t^*(SE_b) = (\_\_\_\_, \_\_\_\_)$

We are 90% confident that for every 1 X variable unit increase, the Y variable increases btw ____ and ____ units on average.

ANALYSIS OF GENDER HEIGHTS

Summary Stats:

(do not copy this table from Fathom. Instead, type these on your slide neatly)

Collection 1

	gender		Row Summary
	F	M	
height	64.5278	69.3561	67.2364
	18	23	41
	62	65.5	62
	63	68	64.5
	64.25	69	67.5
	66	71.5	69
	68.5	73.5	73.5
	1.85085	2.26593	3.18825

S1 = mean ( )
S2 = count ( )
S3 = min ( )
S4 = Q1 ( )
S5 = median ( )
S6 = Q3 ( )
S7 = max ( )
S8 = s ( )

2 SAMPLE T TEST ON THE MEAN HEIGHT OF MALE/FEMALE

Ho: μ males = μ females
Ha: μ males ≠ μ females

Conditions & statement

$$t = \frac{\text{mean1} - \text{mean2}}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}} =$$

2* P(t > _____| df = ___) =

- We reject…..
- We have sufficient evidence ….

CONFIDENCE INTERVAL

- If you reject Ho, complete an appropriate confidence interval

ANALYSIS OF ARM LENGTH BY GENDER

Collection 1

	Gender		Row Summary
	F	M	
Arm_Length	19	23	42
	26.2632	28.587	27.5357
	1.51262	1.50493	1.89477
	24	26	24
	25	27.5	26
	26	28	28
	28	30	29
	29.5	31	31

S1 = count ( )
S2 = mean ( )
S3 = s ( )
S4 = min ( )
S5 = Q1 ( )
S6 = median ( )
S7 = Q3 ( )
S8 = max ( )

DO NOT put this type of table on your power point!! Write out the numbers neatly.

2 SAMPLE T TEST ON THE MEAN ARM LENGTH OF MALE/FEMALE

Ho: μ males = μ females
Ha: μ males ≠ μ females

Conditions & statement

$$t = \frac{\text{mean1} - \text{mean2}}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}} =$$

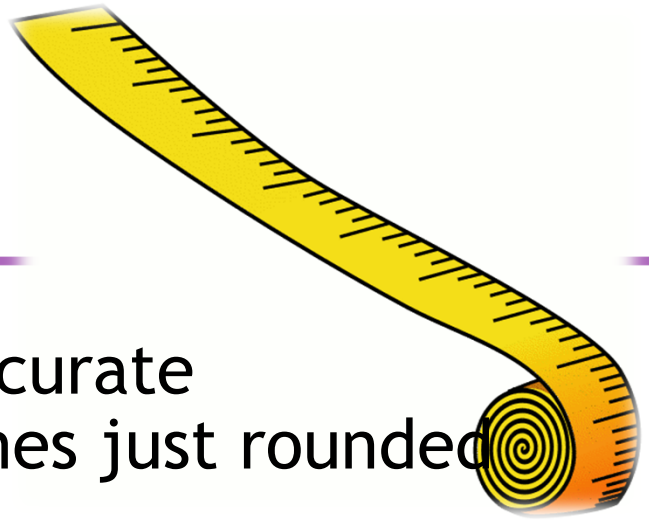
2* P(t > _____| df = ___) =

- We reject…..
- We have sufficient evidence ….

CONFIDENCE INTERVAL

- If you reject Ho, complete an appropriate confidence interval

BIAS AND ERRORS



- We did not take the most accurate measurements and often times just rounded to the closest half of inch
- With people who were wearing pants it was sometimes hard to tell where the ankle was when measuring the lower leg
- People who had their hair styled up higher than their head made it hard to accurately measure their height
- Sometimes people didn't hold out their arms perfectly straight and that could have possibly affected the measurement

CONCLUSION

- Lower leg was the best predictor
  - Strong correlation, linear original plot, scattered residual plot
  - Resistant to gender
- Males had a slightly higher correlation overall
- Most of the time females had slightly lower measurements
- Arm length had the highest correlation (.835)
  - Wrist circumference had the lowest (.637)
- Future
  - More specific measurements
  - More consistent with requirements